The ODUM INSTITUTE
FOR RESEARCH IN SOCIAL SCIENCE

Getting Your Feet Wet with Stata
SESSION ONE
2020

Instructor: Cathy Zimmer
919-962-0516, cathy_zimmer@unc.edu

1) INTRODUCTION

a) Who am I? Who are you?

b) Overview of Course
   i) Exploring Stata
   ii) Working with Data Saved in Stata
   iii) Working with Stata Syntax
   iv) Putting Two or More Stata Data Files Together
   v) Reshaping Stata Data Files
   vi) Analyses and Output

c) Bring and use flash drives!

d) Bold face will indicate a Stata command.

e) For more information about Stata – go to www.stata.com on the web.

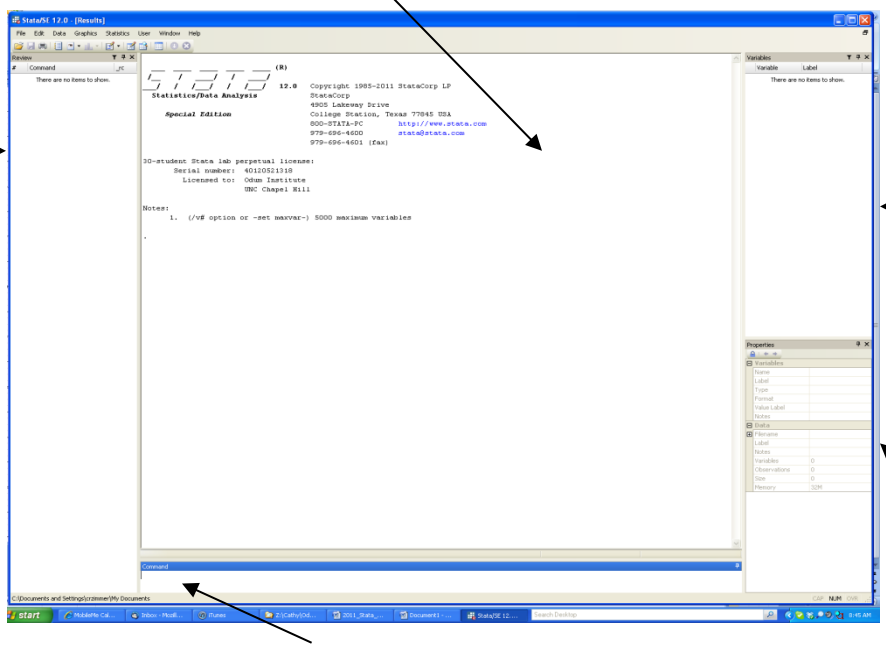
2) EXPLORING Stata

a) Double click Stata icon to open the program. 

b) There are usually five windows to use – adjust them with your mouse.

Results window: results are displayed here.

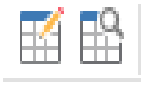
Review Window: past commands appear here -- click on command in the review window, and it will appear in the command window – double click on command here and it will execute.



Variables window: variable list appears here. Double click on variable and it will appear in the command window.

Properties window: variable properties and dataset properties appear here.

Command window: commands are typed here -- the general format is *<command> , <option(s)>* To execute a command, press enter.

c) Opening a Stata data file.
   i) Click File
      Open or click the open file button. 
      Find the file **I:\Cathy Zimmer\STATA\auto.dta**
   ii) Double click on **auto** – this brings the data into another window,
   you can see it by clicking on the Data Editor button or the Data Browser button. 
   iii) OR type in the command window:
      use " I:\Cathy Zimmer\STATA\auto.dta", clear
   iv) Browser can look at specific observations or variables.
      br make mpg foreign
      br mpg-turn
      br in 15
      br in 23/27
      br if mpg==15
      Press enter after each command.

d) Go back to the beginning windows – note that the commands you execute show in the Review window and the Results window. The variables appear in the Variables window.

e) Checking the contents of your data.
   i) Type in the command window: **describe**
      Press enter.
   ii) Describe can be limited to particular variables: **d make price**
      Press enter.
   AND
   iii) Type in the command window: **codebook**
      Press enter.
   iv) Codebook can be limited to particular variables: **codebook make price**
      Press enter.

f) Changing storage type for efficiency.
   i) Type in the command window: **compress**
      Press enter.
   ii) Note changes to storage types.

g) Saving data so you can practice with it. 
   i) Click File
      Save As… or click diskette button.
   ii) Type in file name, keeping ".dta". Choose location for file. Click Save.

h) Saving output on so you have a record of your analyses. 
   i) Open a log file by clicking on the log button.
   ii) Type in file name, replacing ".smcl" with ".log". Choose location for file. Click Save.

i) Viewing the log.
   i) Click on log button.
   ii) Click button next to View snapshot of log file, then click OK.

j) Stata has wonderful help functions – click on Help.

k) To stop Stata at any time from running commands, click on the X button. 

3) WORKING WITH DATA SAVED IN Stata

a) Listing data.
Type in the Command window: **list price mpg foreign**
Press enter.

b) Renaming variables.
Type in the Command window: **rename make car_name**
Press enter.

c) Labeling variables.
Type in the Command window: **label variable car_name "Car Name"**
Press enter.

d) Labeling values.
Type in the Command window: **label define yesno 1 "Yes" 0 "No"**
label value foreign yesno
Press enter.

****Note: you can use the Variables Manager for these commands as well. Let's look at it. 

e) Reordering variables for viewing.
Type in the Command window: **order foreign mpg price**
Press enter.
Another example: **order _all, alpha**
Check out other options for reordering as well.

f) Sorting variables.
i) For ascending sorts only.
Type in the Command window: **sort mpg**
Press enter. Browse data to see sorting.
ii) For ascending and descending sorts.
Type in the Command window: **gsort +foreign –rep78, mfirst**
Press enter. Browse data to see sorting.

g) Producing frequency distributions – the tabulate command.
Type in the Command window: **tab1 foreign**
Press enter.
To see the same frequency distribution with the numeric values instead of with value labels, type in the Command window: **tab1 foreign, nolabel,** Press enter.
To produce frequency distributions for several variables, type in the Command Window: **tab1 foreign mpg car_name,** Press enter.
To produce frequency distributions including missing data, type in the Command Window: **tab1 rep78, missing,** Press enter OR **tab1 rep78, m**

h) Producing histograms.

Type in the Command window: **histogram mpg**

Press enter.

To save a graph, click on File, Save As, type in name for file, and save.

Notice the .gph extension, which tells Stata that this is a graph file.

i) Computing and changing variables.

The generate and replace commands allow you to create new variables or change the values of existing variables.

gen lengthsq=length*length
gen price1000=price/1000

gen rep78r=rep78
replace rep78r=0 if rep78==.

gen exp=0
replace exp=1 if price1000>=10

j) Dropping and keeping variables and observations.

The commands 'drop' and 'keep' are used to drop and keep, respectively, both observations and variables.

drop trunk
drop in 7
drop in 23/27
drop if mpg==12
keep turn

Browse the data to see dropping and keeping.

k) Recoding.

There are many ways to recode variables. Examples of how recodes can be done are listed here, but they are not applicable to the sample data. The x is a variable name and the numbers are its codes.

recode x 1=2
recode x 1=2 3=4
recode x 1=2 2=1
recode x 1=2 2=1 * =3 (* means all other values)
recode x 1/5=2
recode x 1 3 4 5=6
recode x 1 3/5=6
recode x 1 3/5=6 2 8=3
recode x 1 3/5=6 2 8=3 *=1
recode x min/5=min (or max)
recode x .=9
recode x 9=.
recode sex 0=2
recode relat 0=1 *=0

Try recoding rep78 into a new variable called rep78r where 1-3 equal 0 and 4-5 equal 1. Run a frequency distribution for rep78 and rep78r to make sure you did it correctly.

l) Working with part of a string variable.

There are many string functions that allow you to work with parts of string variables.

gen smake=substr(make, 1,3)

m) Converting string variables to numeric ones and vice versa.

i) Numeric to string.

Type in the command window: **decode foreign, gen (sforeign)**

OR

Type in the command window: **gen slength=string(length)**

ii) String to numeric.

Type in the command window: **encode sforeign, gen (nforeign)**

OR

Type in the command window: **destring slength, gen (nlength)**

EXERCISES: Using data from the 2018 General Social Survey,

1. Open the data set (**Z:\Cathy Zimmer\STATA\2018 u.s. general social survey.dta**).
2. How many cases are there? How many variables?
3. Produce a frequency distribution for the **happy** variable for females only. How many cases are missing data?
4. Calculate a new variable that is the average of mother's and father's education. Call it **avged**.
5. Produce a histogram for the new variable. How many cases are missing data?
6. Run a frequency distribution for number of siblings (**sibs**). Then recode into a new variable called **catsibs** with FOUR categories: none, one, two, three or more. Check your recoding by running a frequency distribution and comparing it to the original one.